Representation and Performance Issues in Navigating Visible Human Datasets

Arthur W. Wetzel[a], Brian Athey[b], Fred Bookstein[b], William Green[b], Alex Ade[b]
and Members of the University of Michigan NGI VH Project Team
[a]Pittsburgh Supercomputing Center *(PSC)*.
[b]University of Michigan *(UMVHP)*.
awetzel@psc.edu, bleu@umich.edu, fred@brainmap.med.umich.edu

Abstract

Large volumetric datasets such as the Visible Human present a number of implementation challenges to producing economical yet high speed unconstrained navigation and visualization. These difficulties are magnified when coupled with a need to support many simultaneous users with independent access patterns by means of remote network services.

The Pittsburgh Supercomputing Center has been working together with the University of Michigan project, "Next Generation Internet Implementation to Serve Visible Human Datasets", to produce a system providing multi-user real time unconstrained navigation over the Visible Human data. The Edgewarp program previously developed by Bookstein and Green had demonstrated this 3D navigation capability for single users over constrained preselected volumes. In this project, Edgewarp is being extended as client server implementation able to navigate the full Visible Human dataset using NGI services with the goal of supporting laboratory access by anatomy students. Due to the size of Visible Human data sets previous systems have been based on disk storage. Performance targets for our project lead to a server implementation which maintains volumetric data in primary computer memory and achieves economy of operation by sharing resources over a large number of users on low cost viewing stations. A key portion of the implementation is the use of compressed volumetric data representations to provide efficient and economical data access.

This paper provides an overview of our project capabilities and introduces some of its strategies for navigation and data delivery.

Keywords: edgewarp, volumetric data compression, wavelet, navigation of medical images

Table of Contents

Introduction

Large volumetric datasets such as the Visible Human present a number of implementation challenges to producing economical yet high speed unconstrained navigation and visualization. These difficulties are magnified when coupled with a need to support many simultaneous users with independent access patterns by means of remote network services. The Pittsburgh Supercomputing Center has been working together with the University of Michigan project, "Next Generation Internet Implementation to Serve Visible Human Datasets", to produce a system providing multi-user real time unconstrained navigation over the Visible Human data using network based client server methods.

The application for our project is the training of students in the anatomy laboratory. In this setting the students are working on actual cadavers but need assistance in locating and identifying anatomical features and understanding their positional relationship to other structures. The Visible Human data, when appropriately labeled and presented to students as they work, will provide a valuable teaching asset and could reduce the amount of help needed from laboratory instructors. Our working model is to bring inexpensive viewing computers into the anatomy laboratory so that every student can access all of the Visible Human data and derived information on demand in real time interactive form.

Unfortunately, PC class machines are not yet powerful enough to economically store and manipulate all of the required data especially when duplicated for every individual user station. The solution to this problem is to hook the viewing computers over a high speed network link back to a more powerful server computer which can provide the shared data storage and additional computing horsepower as needed. To support the target goal of 40 simultaneous users there must be a careful balance between client side and server side responsibilities and a judicious use of network bandwidth.

Currently, as we approach the end of year one of our three year implementation project, we are testing the initial versions of our user interfaces, server support and network delivery codes. The body of this paper will review some of the capabilities we are building into this system and present some of the tradeoffs in representation and implementation methods that are being used to build the system.

System Overview

This work builds on a history of visible human browsing software developed by Dr. Brian Athey's team at the University of Michigan and implemented by Alex Ade. ([1]) ([2]) These early browsers operated from a collection of preconstructed orthogonal slice images representing specific anatomical regions. Figure 1 shows screen captures from the head and pelvis browsers. The user interface provides for slice selection in any of the cardinal planes and at three standard resolutions. The extension of this paradigm to the system currently under development provides access to the entire body as shown in Figure 2. In addition to the normal RGB cryosection images the user can also retrieve CT and MRI image data.

Facilities have recently been added to link the image domain to a database of anatomical labelings and terminology as shown in Figure 3. Using this mechanism, a student can select a feature by name and its location and bounding box is highlighted in the image panels. The inverse mechanism, shown in Figure 4 , lets the user point to a region in the image to retrieve specific text data on that anatomical feature. Of course most anatomical features are not aligned with the orthogonal data axes. As shown in Figure 5 and Figure 6 even a small rotational offset from the alignment of an anatomical feature can make a great difference in visibility.

The basis for many of the new capabilities which distinguish the system we are developing from the previous static orthogonal slice browsers comes from the Edgewarp software developed by Drs. Fred Bookstein and William Green. Edgewarp was built to support comparative studies of shape differences and their associations with biological causes and effects. This area of study has been named morphometrics. Edgewarp was initially implemented for for use with two dimensional shapes and was later extended for use with 3D volumetric data in the context of human brain studies. A central feature of Edgewarp is its use of thin plate deformable splines to model shape differences and the use of bending energy as a measure of difference between shapes. ([3]) ([4])

In addition to its application for morphometrics, Edgewarp-3D has demonstrated a powerful capability for navigation and visualization within volumetric medical data sets by the display of arbitrarily oriented cutting planes interactively with real time continuous motion. In the short term for this Visible Human application we are primarily concerned with this visualization aspect rather than the underlying shape comparison mechanisms. In the future, as more examples of Visible Human data sets are produced, we expect morphometric functions of Edgewarp will be very useful.

At the beginning of the current project Edgewarp was already implemented for manipulating single channel volumetric data in the form of a self contained data cube. This cube of data was read into computer memory and slices displayed as gray scale moving images as the user navigated through the volume. Recently this capability has been extended for use with 3 channel RGB color data which produces an interactive user interface as shown in Figure 7. In this example, from the sholder region of the Visible Female data, the "scout" window on the left provides visual context about the location of the current slice plane and the right window contains the direct view of that plane. These capabilities are described in the Edgewarp user manual, available from *ftp://brainmap.med.umich.edu/pub/edgewarp3.1/*, and in a recent publication. ([4])

Representation and Performance Issues

The operation of Edgewarp for visualization is most effective when running on a platform that is fast enough to provide the illusion of continuous motion through the data. In this mode, both window panes from Figure 7. are dynamically updated at video rates. The dynamic real time behavior is especially important for fine scale positioning and visual feature location tasks which become much more difficult and tedious if the update rate falls below about 3 frames per second. Until recently this speed required an expensive high end visualization computer such as the Silicon Graphics Onyx. However, improvements in personal computers and inexpensive graphics cards, such as the Nvidea models, now provide an appropriate platform for less than $2000.

In this project, Edgewarp is being extended beyond the constrained volume and single user implementation to operate in a multi-user client server mode. This will allow many users to simultaneously navigate the full Visible Human dataset using NGI services to exchange volume and image data between inexpensive viewing stations and a much more powerful server system which maintains the full data store. The greatest technical challenge is to maintain the essential dynamic and continuous Edgewarp behavior while operating with remotely fetched data.

Further evolution of the Edgewarp style framework is illustrated in prototype form in Figure 8 and Figure 9. In these examples the control button interface is simplified for use by nonexpert users who do not need the additional special purpose controls. Also, the use of transparent overlays provides additional context so that users have a better visual impression of their location and orientation in the dataset.

General purpose networked data delivery from the server has also allowed us to prototype the mechanisms shown in Figure 10 and Figure 11. Since many anatomical features have curved pathways along with some natural extent that can be mapped to a plane, it is useful to be able to construct flat images from curved data surfaces. In the illustrated example, the user has selected points along the spine from a planar side view which have been connected by a spline fit. By projecting image scan lines perpendicular to the spline path we reconstruct a flattened face on view of the spinal column which reveals features that can not be seen over a large extent in any single planar view.

Due to the size of the Visible Human data sets previous systems for delivering non axis aligned views have been based on disk storage. Some of the capabilities of Edgewarp for generating arbitrary cutting planes are also present in the excellent web accessible disk based service from Switzerland provided by Dr. Roger Hersch and his collaborators at the web site *http://visiblehuman.epfl.ch/* and described in several publications including presentations at this conference. ([5]). ([6]). ([7]).

Multi-user performance targets for the Edgewarp style 3D volume navigation capability in our project has lead to a server implementation which maintains the entire volumetric data in primary computer memory at the server and a much smaller working set at the client. As the users region of interest moves through the volume, unneeded portions of the working set are thrown away and new regions are loaded on demand from the server. We will be applying an anticipatory prefetch mechanism during the next year’s work to further reduce effects of retrieval latency. Economy of operation is achieved by sharing the server resource over a large number of users working from low cost client viewing stations. Clearly the cost of storing data in memory is higher than the cost of storing the same data on disk. However, when generating arbitrarily aligned, positioned, and scaled slice images from a disk based data set disk seek time quickly becomes the performance limiting factor.

The seeking bottleneck can be substantially reduced by using many disks in a RAID configuration so they collectively maintain high performance by doing simultaneous seeks as done by Hersch and his collegues. This approach becomes less attractive as disk technology continues to provide rapidly increasing storage capacity per spindle but with seek (ie. data access time) improving more slowly. It is important to note that improved single disk bandwidth also does not greatly improve the type of scattered access needed to assemble arbitrary cutting slices from any static arrangement of data on disk. In particular, if the data is stored in the original slice data capture order as provided from NLM, then the reconstruction of a vertical sagittal slice involves every slice.

This situation is greatly improved in the general case by reorganizing the data to a more direction neutral order which unfortunately sacrifices some of the fortuitous efficiency of horizontal slice retrieval. We have used this principle during our prototyping stage to produce a disk based arbitrary slice server running from disk. The data is reorganized into a set of 64*64*64 cubes which we have called "cubelets". The Hersch group from Switzerland uses the term "extents" for the same concept. The cubelet (or extent) organization brings data that is close together in a 3D region close together on the physical disk. Each cubelet of RGB data occupies 768Kbytes which is a good data size for a single disk read operation. In extracting slice data from cubelets, only a small portion of the data in any single cubelet, 12Kbytes in the case of a cardinal slice, is used. However, in the Edgewarp navigational scenario, it is very likely that much of the additional cubelet data will be used as the user moves through the local volume. Therefore a single disk operation is amortized over a period of operation when the user is working in that same region. Any single extended slice view is constructed by the intersection of the cutting plane with a constrained set of cubelets. However, the number of disk operations is still large enough that it becomes difficult to support more than a few simultaneous users. It is the need to support many simultaneous users in a near real time interactive work mode with arbitrary slice viewing which forces us to a memory based server rather than a service operating directly from disk storage.

Preliminary results of networked Edgewarp operating with an uncompressed data flow show adequate real-time behavior at data network speeds around 2.5Mbits/sec. In this experiment data is requested and delivered in very small "microcubes" of 8*8*8 voxels and at variable resolutions which match the users current scale factor. (At the time of the last revision we have renamed these units "chads" in honor of the election results). For the 3 channel RGB data this means that each delivered unit is 1536 bytes. These microcubes are individually requested. The simple test protocol is a UNIX socket interface between UNIX or Linux versions of Edgewarp back to the server running on a Compaq ES-40 at the PSC machine room in Monroeville Pennsylvania.

In scaling to many simultaneous users, it is important to keep the data in memory to avoid disk delays. A key portion of the memory based implementation is the use of compressed volumetric data representations to provide efficient and economical data access at performance levels adequate maintain smooth user interaction. We can estimate the compression requirement from several respects. First, the raw Visible Human data including both the male and female along with CT and MRI approaches 60Gbytes and the recent release of 70mm film digitizations further increases that. However, only about 20% of the data is body tissue which still leaves 10 to 15Gbytes of anatomical volume. This is too much to keep entirely in memory at a good price point at the current time. However, it is not unreasonable to dedicate 2 to 3Gbytes of server memory for this purpose so we do not need an extreme compression factor to implement the memory based service.

Based on uncompressed networked Edgewarp results we expect to service 40 users with as little as 100Mbytes/sec of network traffic. To set an upper bound however we should note that the raw update rate to 40 users with dual 512*512 window panels, corresponding to the Edgewarp scout and cutting plane views, and updating at 10 frames per second is ~5Gbits/sec which is well beyond even the Gigabit network capacity especially when one considers the usual 40% network loading rule of thumb. The 50:1 discrepancy between these two bounds is the result of a number of factors. The most important of these are the downsampled and interpolated Edgewarp view that is produced during dynamic navigation and the reuse of the client side working set data cache during normal operation. The downsampled view during dynamic navigation takes advantage of the visual perception characteristic which makes image detail much less important during motion. Therefore we can fill in a full resolution display when the user stops and get by with a lower quality display the rest of the time. Additionally, the Edgewarp scout window is constructed from the same data as used for the right window plus static orientation planes so it does not have to be transmitted separately. Despite these factors which tend to conserve network bandwidth and reduce the effects of network latency, we would like to make the best use of bandwidth and other server side shared resources. Data compression is the main way to do this.

We have been evaluating a number of types of compression and the ways in which they help with providing good user service. There are many methods of data and image compression. JPEG, based on the Discrete Cosine Transform (DCT) is the most frequently used for network based image delivery. However, JPEG does present blocking artifacts at high compression ratios and does not take advantage of the third dimension for use with volumetric data as we encounter with this Visible Human application.

There are several specific characteristics which lead us to an alternative wavelet based compression approach which bears similarity to the 2D wavelet basis of the emerging JPEG2000 standard. One of these is the fact that the data is naturally represented in a hierarchical fashion where new details emerge at higher resolution levels but broad area features are largely predicted from low resolution interpolated data. This fits well with the fact that users will frequently change their scale of view to search for a location in broad context and then zoom in for a close inspection of detail.

In recent years, this natural hierarchical encoding property of wavelet transforms has been well studied and reported for video and image applications. ([8]) ([9]) Some recent variations of the general wavelet family are also being applied to surface representations of objects which will be applicable to the segmented tissues and organs from the Visible Human data. ([10]) ([11]) A core group of researchers, Sweldens, Schroeder, Daubechies and Kovacevic has been working with the lifting scheme for wavelet construction which was introduced by Sweldens at Bell Labs. ([12]) ([13]) ([14]) ([15])

The essential concept of lifting is to predict a higher resolution version of a signal, usually a multiple of two, from a lower resolution version and to split the result into the correct portion of the prediction and the error of prediction. The correct portion of the prediction, termed the low pass portion, carries no new information and does not have to be encoded in a compressed representation. The error of prediction, the high pass signal, does carry new information and therefore must be encoded to reconstruct the high resolution signal from a previous low resolution version. This is the basis of the method we are implementing for the in memory storage and network delivery of data. On the server side, a hierarchy of levels of resolution for the Visible Human data is maintained with each new level of resolution encoded as the difference from the previous resolution.

In principle this recursive hierarchy could be reduced back to a single voxel as the root of the representation with the encoding of differences forming a complete tree. Unfortunately, such a method would require the reconstruction of the entire tree from the root in order to access any individual piece of data requested by the user. We need a way to randomly access any portion of data quickly and without a lot of reconstruction.

The representation we are currently implementing uses a base level of representation which corresponds to the 64*64*64 cubelets which is fully indexed and directly addressable. Within each cubelet, a small tree is produced which provides relatively quick access down to the level of 8*8*8 microcubes. Maintaining these levels as directly accessible has the overhead of pointers to the location of each piece of the representation. This is not a problem at high levels (low resolutions) of the hierarchy but is not acceptable at the highest resolutions. Instead, the microcube level is being completely represented as a variable length entropy coded bit string. This means that individual voxels can not be directly retrieved without at least partially decoding the microcube.

In order to take advantage of direct accessibility to the microcube level, each microcube must be completely self contained. This is assisted by encoding a context at the start of each microcube which indicates its general characteristic in terms of data content. This characteristic is a type of data segmentation based on the location of the data in color space and its statistical properties. This is illustrated in part in Figure 12 which shows the very nonuniform distribution of voxel values in the RGB color space from three different vantage points round the RGB color cube. In the corner view of image (a) we see that most of the data clusters along the black to white axis. This is reinforced in views from the side of the color cube in (b) and (c). Also apparent, especially in images (b) and (c), we see that there are groups of voxels with a similar slope in color space. This segmentation mostly corresponds to the anatomy but does not have to be perfect and does not have to distinguish anatomically separate features that would be important to anatomists. Rather, the purpose is to get statistical groupings which provide good predictive values for compressed encodings. Even if the segmentation is not the same as would be produced by an anatomist, the full image detail can be reconstructed by the coding method.

Representation of the average location in RGB space and also the average slope of the voxel values for each microcube gives an excellent and highly constrained space for applying the entropy coding. Each specific region (ie. context) has its own distribution characteristics and tends to lump together a type of tissue. This aspect is shown in Figure 13 image (b) where regions of solid color, as is particularly apparent for muscle in this case, group together. Each of these solid categories, which are generally located together, has its own optimal entropy coding table.

Reconstruction from the encoding is based on differences from the average of that category. This encoding can be lossless which will provide an exact reconstruction of the original full resolution data or lossy which produces a visually close approximation. The lossless coding is limited to a compression ratio of about 3:1 by the inherent signal to noise ratio (SNR) of the data which appears to be about 80:1 in the case of the Visible Female RGB data. Lossy coding allows a potentially much higher compression ratio but has the problem that we must decide on a suitable level of accuracy for the reconstructed rendition. At the current time we are preparing tests sets for visual evaluation by the members of the project to determine these settings and hence the ultimate compression level that we will be able to use in practice.

Conclusions

The requirement to support 40 simultaneous student users accessing Visible Human based training materials from the anatomy laboratory involves many types of networked data delivery. The most technically demanding of these is interactive navigation through the Visible Human volumetric data with associated high performance display of arbitrarily oriented, positioned and scaled cutting planes to simultaneous but independent users as provided by the networked implementation of Edgewarp-3D.

The best balance of server and client responsibilities to support 3D visualization and navigation from a central data server is a moving target as the capabilities and costs of computing and graphics hardware evolve. Within the last year there has been a dramatic increase in the power of low cost PC based graphics, driven by the gaming market. This has allowed us to move away from the graphics supercomputer model (ie. SGI Onyx), where display images would be produced at the server and simply passed to a simple display client, to a more distributed solution where most of the visualization function is handled at the client. This also changes the characteristics of network traffic to support the application and greatly reduces user interaction problems due to network latency.

Even with the falling cost of PC hardware it is still most economical to maintain the current Visible Human data and its supporting database centrally. This model is scalable to larger numbers of users replicating servers when the number of clients exceeds the capacity of a single server. To maximize performance and the number of clients supported by a single server it is most efficient to use volumetric data compression so the server can keep all data in primary memory and also economize on network bandwidth by delivering the data in compressed form when possible. The server is still able to produce completed images if needed to support very low end clients which do not require the level of performance demanded by the primary anatomy laboratory application.

We look forward to reporting the results of user and scalability studies in a future publication as our system becomes operational during the coming year. Additional information, including links to the Edgewarp software and partner sites, can be found at our main project web site *http://vhp.med.umich.edu/*.

Acknowledgement

We gratefully acknowledge the support of NLM contracts N01-LM-8-5630 and N01-LM-0-3511, Brian D. Athey, P.I., which makes our work possible.

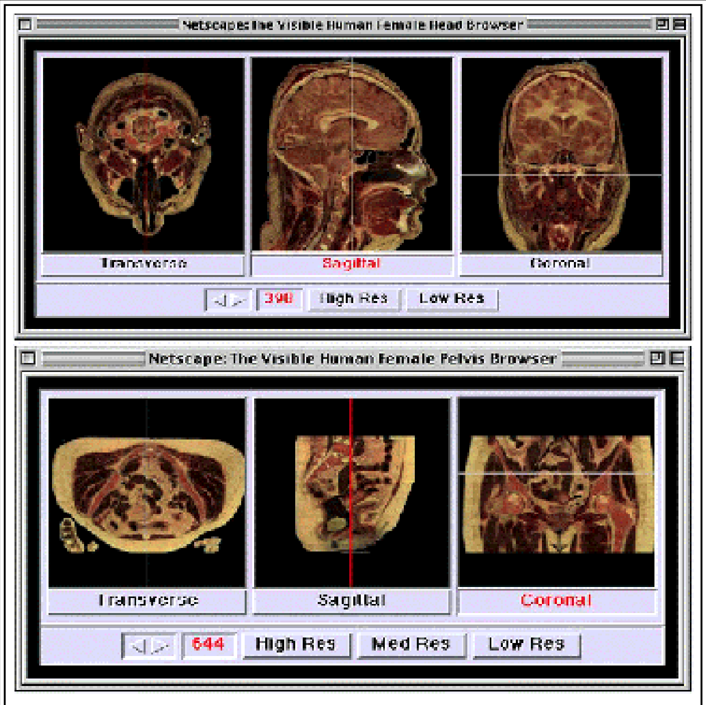

Figure 1. First generation Michigan browsers provided region specific orthogonal slice viewing.

Figure 2. Browsing capability has been extended to cover the entire body and multiple data sets.

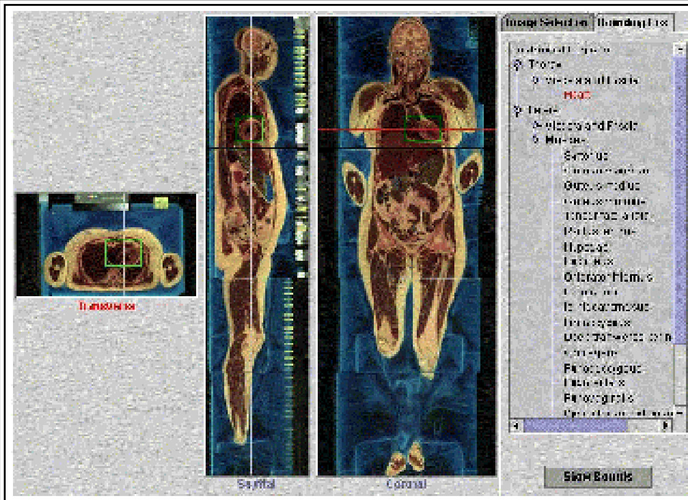

Figure 3. The location and naming of anatomical features is provided by linking the viewing position to a reference database.

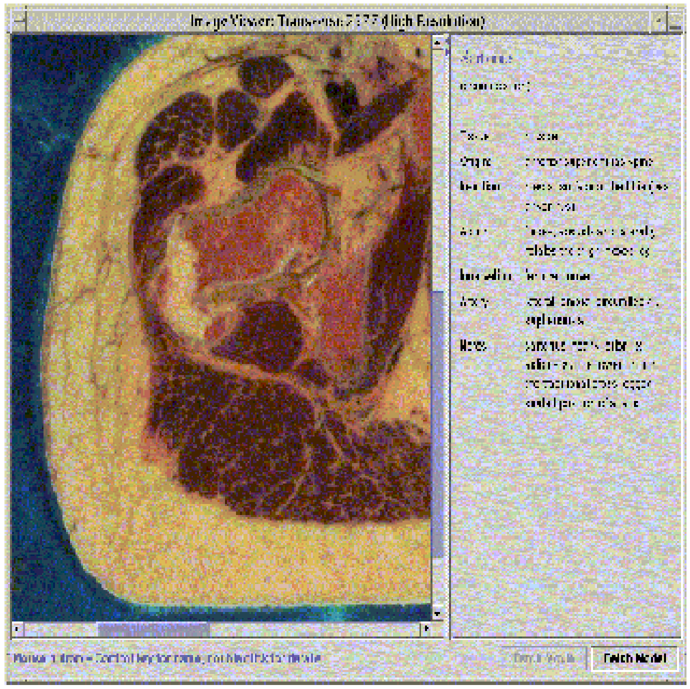

Figure 4. Additional information on specific areas of anatomy are being inserted into the database for user instruction.

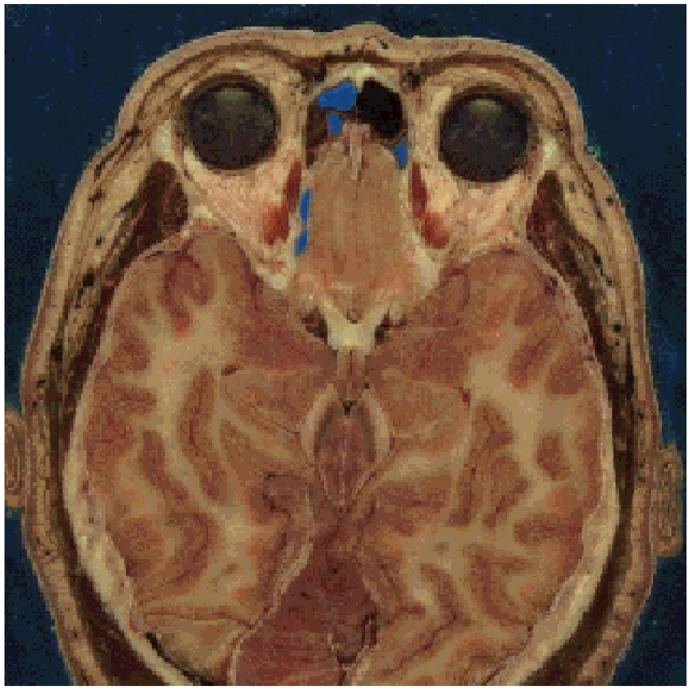

Figure 5. The normal orthogonal image orientations provide only a partial view of important features even if they are only slightly tilted with respect to the image plane.

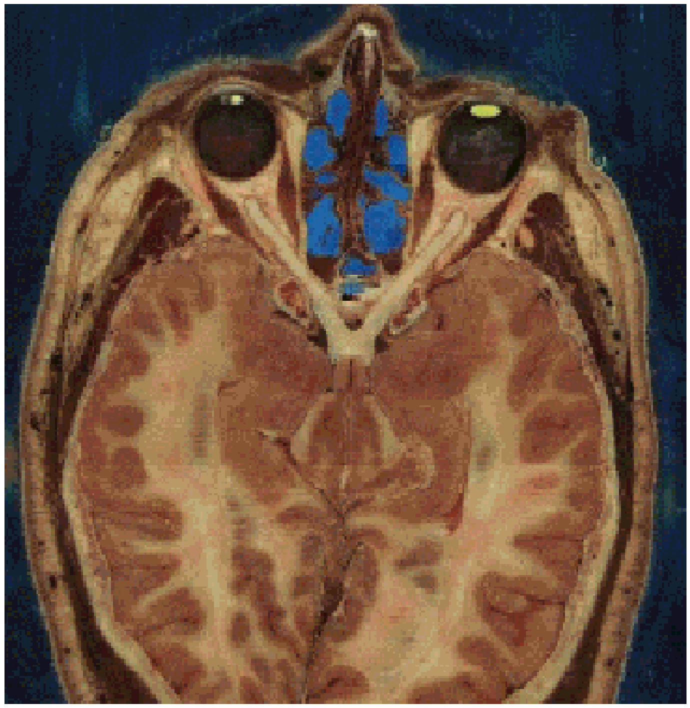

Figure 6. Capability for arbitrary orientation of cutting planes provides access to important relationships in a single planar image.

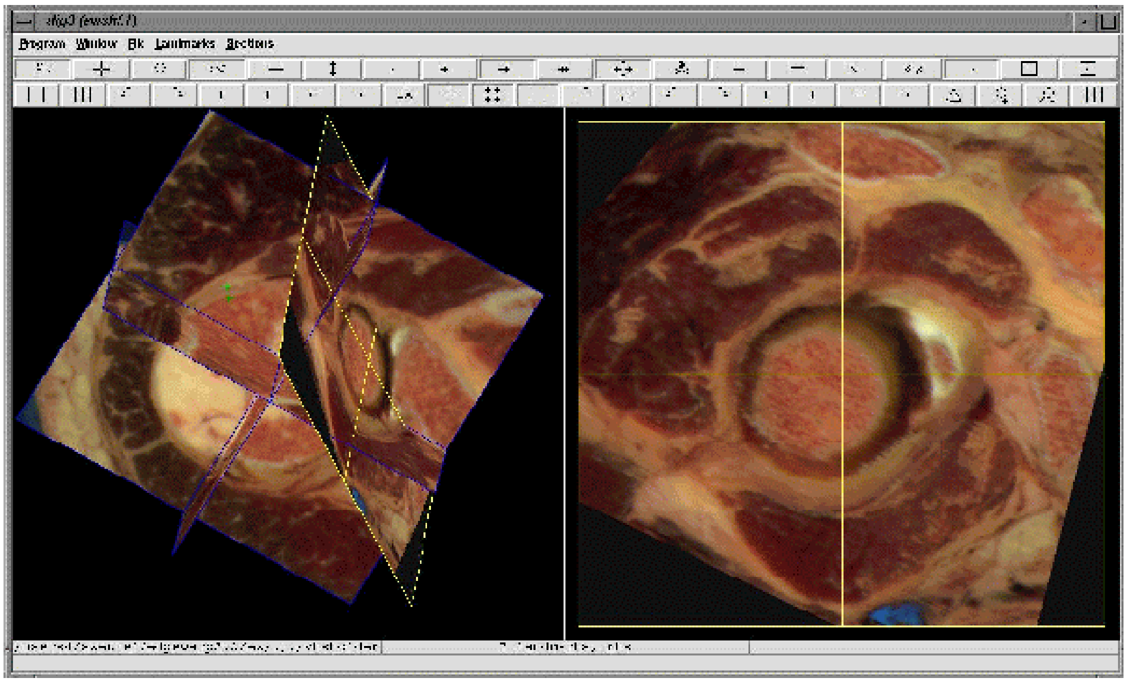

Figure 07. The Edgewarp program provides a powerful facility for producing arbitrary axis slice views and dynamic interactive navigation through Visible Human data.

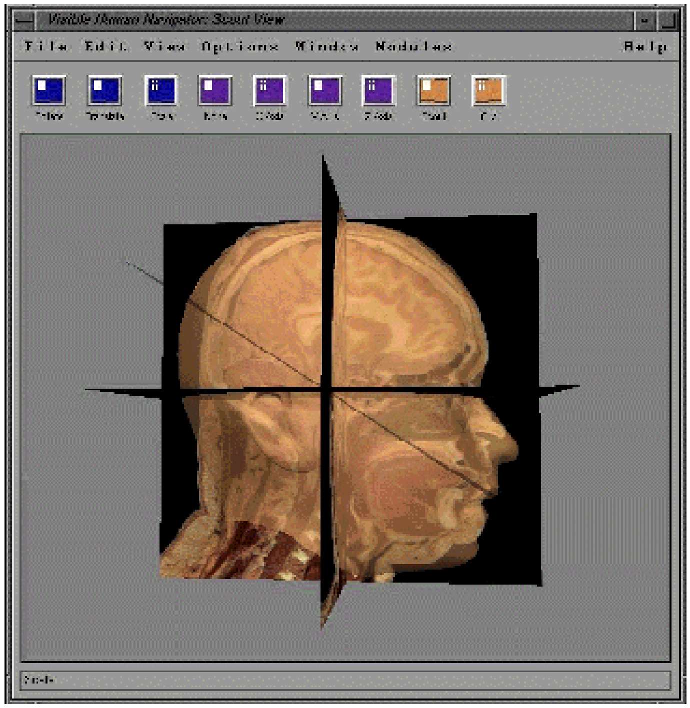

Figure 8. Transparent overlays provide a useful aid to users for understanding anatomical orientations.

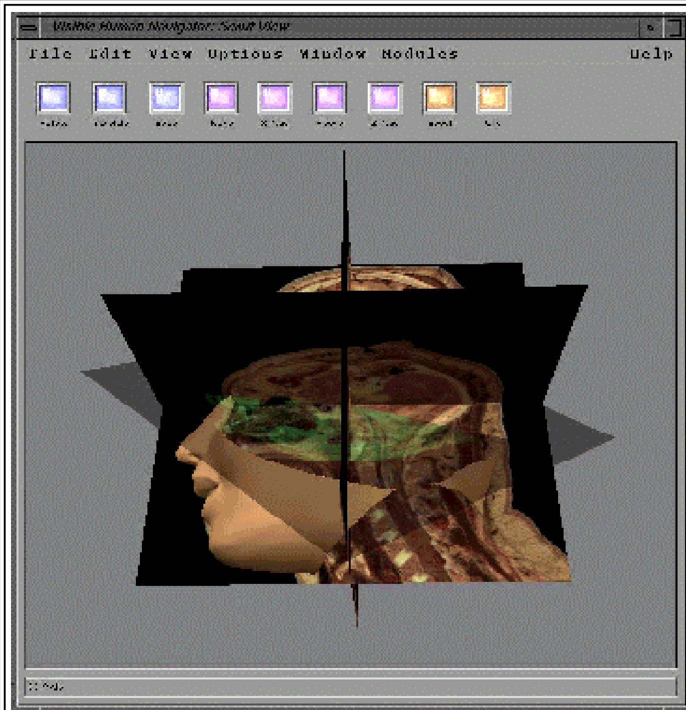

Figure 9. The addition of context lets users accurately position cutting planes to locate features of interest.

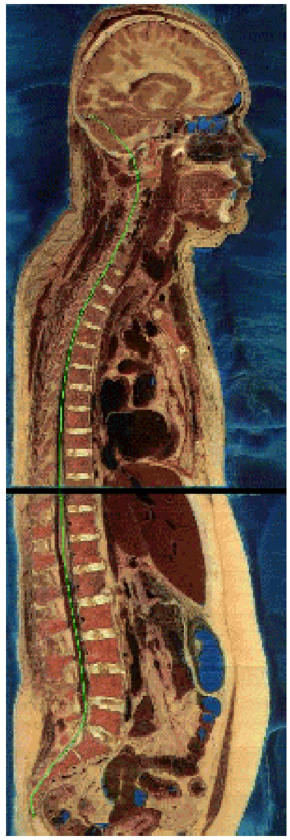

Figure 10. Curved pathways such as the spine can not be adequately captured by any single cutting plane.

Figure 11. Mapping a curved surface following the spinal pathway onto a flattened image lets us see the entire spinal column in a single view.

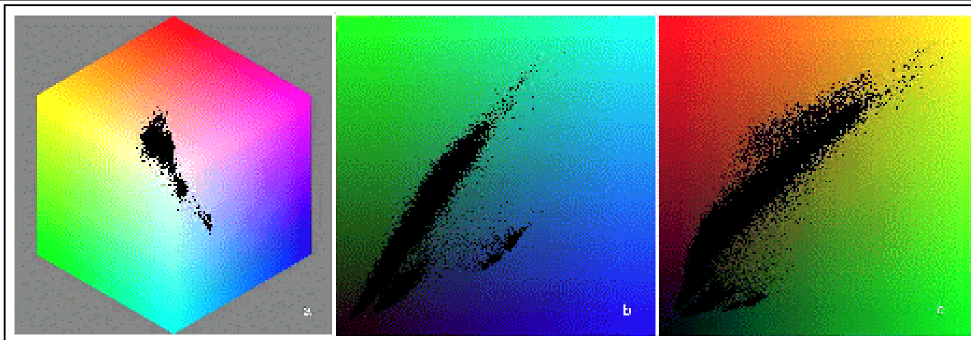

Figure 12. Clustering in the color distributions corresponds primarily to tissue categories, such as fat or muscle, and provides a basis for efficient compression. These three images show the distribution within the RGB color cube from three different cube orientations.

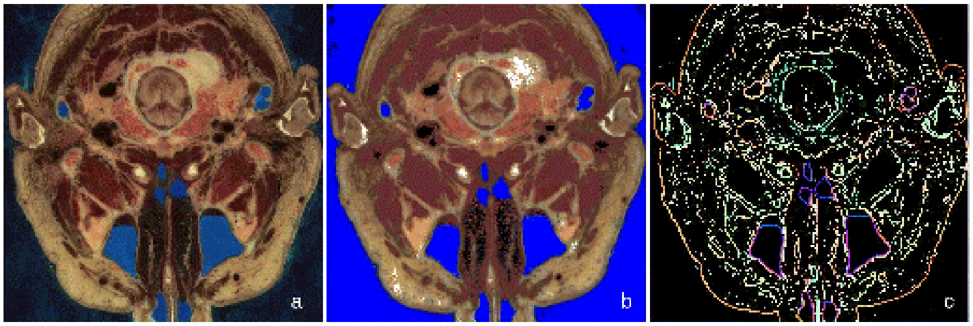

Figure 13. Color clusters from image (a) lead to a reduced color mapping in (b). Edge detection in image space predicts relative compression effectiveness. Flat areas, which appear dark in image (c) compress very well while edge areas produce a larger compressed code.

References

[1] B.D. Athey, A.J. Warner, J.C. Laby, W.M. Meixner, J. Chung, J.P. Williams, P.V. Ketty, Using the Visible Human as an image database locator, The Visible Human Project Conference Proceedings 1996 Oct 7-8

[2] A. Ade, W. Meixner, B. Athey, The Visible Human Female World Wide Web browser. 2nd Visible Human Project Conference Proceedings; 1998 Oct 1-2; Bethesda, MD

[3] F.L. Bookstein, W.D.K. Green, Edgewarp: A flexible program package for biometric image warping in two dimensions, Proceedings of SPIE Vol. 2359 (1994) pp 135-147

[4] F.L. Bookstein, B.D. Athey, W.K. Green, A.W. Wetzel, Navigating solid medical images by pencils of sectioning planes, Proceedings of SPIE Vol. 4121 (2000) pp 117-127

[5] S. Gerlach and R.D. Hersch, The Real-Time Interactive Visible Human Navigator, Third Visible Human Project Conference Proceedings, Oct 2000

[6] J.C. Bessaud and R.D. Hersch, The Visible Human Slice Sequence Animation Web Server, Third Visible Human Project Conference Proceedings, Oct 2000

[7] S. Vetsch, V. Messerli, O. Figueiredo, B, Gennard and R.D. Hersch, A Parallel PC-Based Visible Human Slice WEB Server, Second Visible Human Project Conference Proceedings, Oct 1998

[8] M. Vetteri and J. Kovacevic Wavelets and Subband Coding, Prentice Hall 1995, pp 443-448

[9] G. Karlsson and M. Verterli, Three-dimensional subband coding of video, Proce. IEEE Int. Cof. Acoust., Speech and Signal Proc., April 1988, pp 1100-1103

[10] I. Guskov, K. Vidimce, W. Sweldens and P. Schroeder, Normal Meshes, Siggraph 2000

[11] A. Khodakovsky, P. Schroeder and W. Sweldens, Progressive Geometry Compression Siggraph 2000

[12] I. Daubechies and W. Sweldens Factoring Wavelet Transforms into Lifting Steps, J. Fourier Anal. Appl., Vol. 4, Nr. 3, pp. 247-269, 1998

[13] W. Sweldens, The Lifting Scheme: A new philosophy in biorthogonal wavelet constructions, Wavelet Applications in Signal and Image Processing III, pp. 68-79, Proc. SPIE 2569, 1995

[14] R.C. Calderbank, I. Daubechies, W. Sweldens and B.L. Yeo, Losless Image Compression using Integer to Integer Wavelet Transforms International Conference on Image Processing (ICIP), Vol. I, pp. 596-59

[15] J. Kovacevic and W. Sweldens Wavelet Families of Increasing Order in Arbitrary Dimensions IEEE Trans. Image Proc., vol. 9, nr. 3, pp. 480-496, 2000